

DH Box

Final Performance Report

The Graduate Center, CUNY, March 2019

Overview	2
Project Activities	2
2015	3
2016	3
2017	4
2018	5
Audiences	5
Evaluation	6
Continuation of the Project	7
Long-Term Impact	7
Award Products	9
Conferences	9
Workshops	9
Publications and Press	10

Overview

In 2014, DH Box was envisioned as a pedagogical tool for DH-inflected classrooms and workshops, a "digital humanities laboratory in the cloud" that would give students the opportunity to engage with research questions without installing, configuring, and troubleshooting a digital working environment. Entering the NEH grant period, the DH Box team had already developed a working prototype and, for an early-stage project, had received a number of offers of collaboration and expressions of interest from a wide variety of scholars and institutions. In the years the DH Box team has worked on the project, we've exceeded our own expectations for what could be accomplished with startup funding from the NEH. We've added new tools, redesigned the interface, and made the platform more stable. We've taught classes and workshops and seen others adopt DH Box for their own pedagogy. Perhaps most importantly, at places like the Digital Humanities Summer Institute and the Modern Language Association conference, we've had the opportunity to engage in discussions about humanities infrastructure, service, inequality, and the role of DH practitioners in the academy.

Even as we celebrate our accomplishments in teaching, writing, programming, and speaking over the past two years, from our vantage at the end of the grant period we feel that there is much yet to be done with DH Box and similar projects dedicated to providing infrastructure by and for scholars and students in the humanities. In 2014, the DH Box project was an early

adopter and advocate for using cloud technology to provide less expensive and more widely available infrastructure. Now, as the cloud and cloud services increasingly dominate in higher education and in industry, and adoption of for-profit cloud services moves toward ubiquity among teachers and students, we feel both that humanities-inflected infrastructure is even more critical and that the challenges facing this kind of intervention are even greater.

In 2015-2016, the DH Box team made the deliberate choice not to provide DH Box as a service. Though we operate a demonstration site that allows users to try out the software before installing it, there is a DIY element to DH Box—users have to install the software on their own cloud servers, not use ones provided by the team. This choice was partly ideological and partly practical, and though we acknowledge that offering a service is not incompatible with an open source, DIY culture, we stand by this decision. Though we think this choice was the right one, we have come to realize that setting up cloud infrastructure, even with a relatively compact process, can often be as difficult for prospective teachers as it is for their students.

Project Activities

In the original grant application, the DH Box team focused on two goals for the project during the grant period: outreach and platform stability. While the team had a working and usable prototype, at the time certain technological bottlenecks made it less practical and usable. These efforts toward platform stability would coincide with encouraging other scholars to become early adopters, giving feedback on the platform as we worked to improve it.

DH Box was first envisioned as a project by a small team of students, including current lead developer Stephen Zweibel, as part of the Graduate Center's Digital Praxis Seminar. During the seminar, the project changed course from a minimal computing model—using \$35 Raspberry Pi computers to set up a local digital humanities lab—to a cloud computing, browser-based model. DH Box would therefore be server-side software, like WordPress or Omeka, that could be installed by users on a cloud service such as Amazon Web Services or DigitalOcean and which could be accessed via browser at a particular domain or IP. The initial version of this software, created in the Digital Praxis Seminar, would use a technology called Vagrant designed to automate the process of setting up a development (programming) environment. Though this approach worked, creating the environment on user request would take a long time—often more than five minutes, which is an eternity in the context of users used to fast responses from web applications. Entering into the grant period, therefore, the team's first goal was to speed up the process of creating these environments to something acceptable to users.

2015

Shortly before September 2015, the beginning of the grant period, the DH Box team learned about the containerization technology called Docker that could create these environments much more readily and which had a number of other technical advantages over Vagrant. As the grant period began, the team was already making the transition, and by winter had a version of DH

Box that used Docker to create "containers"—essentially lightweight virtual machines—in under one minute. In testing at workshops at the Graduate Center and in a class at Hunter College, this time was much more acceptable to students and other users of the platform. Also during this first semester of the grant period, the DH Box web application underwent an audit by an external programmer and a number of browser-specific issues were found and corrected through feedback from alpha testers. Finally, a UX specialist at the Graduate Center evaluated the platform, resulting in a new UI design with increased usability and greater accessibility for disabled visitors.

During this time, a demonstration site for DH Box went live, allowing potential users to test the platform before choosing to install it. This demo site also served as the primary site for testing DH Box in workshops at CUNY. This initial period also saw the beginnings of a collaboration with Jonathan Reeve, a graduate student at Columbia who began developing a corpus downloader for use on the platform.

2016

From January 2016 onwards, the DH Box team focused on adding frequently requested tools to the platform, including Brackets, a text editor, WordPress, and IPython (now Jupyter Notebooks). Attention was also given to the command line emulator, and the demonstration site was moved to a larger Amazon Web Services instance to facilitate testing of more computationally demanding tasks in workshops.

One well-received addition during this time was the development of a tabbed browsing interface, one that enabled users to better understand that DH Box was not comprised of siloed applications but consisted of multiple interfaces to the same containerized computer. Rather than open new browser tabs for each DH Box service (such as Brackets, the command line, or Jupyter Notebooks), users could switch among these contexts using tabs native to DH Box. One of the tabs implemented in this feature was an open port (port 5000), so that users running web servers associated with tools such as Jekyll or Flask could view these running processes within DH Box.

During this period, the DH Box team presented at the Canadian Society for Digital Humanities at the 2016 Congress in Calgary, Alberta and tested DH Box through a series of introductory Python workshops as part of the Interactive Technology and Pedagogy program at the Graduate Center.

2017

In 2017, the DH Box team refocused from platform usability and adding features to the long-term sustainability of the project. This included resolving security issues, creating documentation, and continuing ongoing outreach efforts.

In January, the DH Box team presented in a session on "Open Source Lit, Open Source Crit" at the Modern Language Association conference in Philadelphia, a lively and well-attended discussion that explored the intersections between text and technology and the role of infrastructure in the humanities. The year saw the adoption of DH Box by Archeology Box, a project headed by Shawn Graham at Carleton University, and also the use of DH Box in a session on minimal computing at Columbia. The DH Box team also entered discussions with NYU's Center for Urban Science and Progress (CUSP) and the British Library, both of which had teams who were considering internal use of DH Box for projects in their organizations. In June, a discussion between team members Stephen Zweibel and Patrick Smyth was published on DH+Lib, titled "DH Box and Access in the Digital Humanities." In August, the DH Box team traveled to Montreal for the Digital Humanities 2017 conference, where they presented a poster on the platform.

In discussions with organizations like the British Library, the DH Box team learned that improved documentation for the project would encourage its wider adoption and improve sustainability. To this end, in 2017 the team worked toward creating a more complete "getting started" guide for DH Box and a technical overview of the platform. DH Box also participated in a Doc-a-Thon at the Graduate Center in which members of the public contributed to open source projects. These efforts have resulted in improved documentation for DH Box, though more could still be done on that front to ensure the sustainability of the project.

An ongoing challenge in 2017 was the security of the platform, an issue which included password and application security as well as the isolation of the generated Docker programming environments from one another and from the underlying system. Though the DH Box team made critical changes in these areas, no security consultant within the budget of the project could be found to provide an external assessment of the project. For many use cases of the platform, these considerations are less important. However, security concerns may have an effect on adoption of the platform, especially when use of DH Box must be approved by local IT.

2018

The 2018 grant extension was requested primarily to allow for maintenance of a public DH Box presence on Amazon Web Service instances. However, this year saw strong activity for DH Box as a project, including the use of the platform by the DH Box team (Stephen Zweibel, Jojo Karlin, Jonathan Reeve, and Patrick Smyth) at the basis of a week-long course on Web APIs with Python at the Digital Humanities Summer Institute in June. The course was taught entirely using DH Box in a computer lab with no administrative access to the local machines. Students were trained in the command line, including tools such as curl, used Brackets to write Python scripts, and employed Jupyter Notebooks to visualize data returned by API. Students also created their own projects, including twitterbots and visualizations, all using DH Box. Based on feedback collected by DHSI, the course was well received, and the DH Box team will lead another similar course in 2019.

In 2018, Jonathan Reeve, a member of the DH Box team, released CorpusDB, an API based on the code initially developed for DH Box to download text corpora for analysis on the platform. CorpusDB allows for accessing versions of texts from sources such as Project Gutenberg in forms ready for text analysis, and also provides metadata cross-referenced from WikiMedia and pre-run distant reading artifacts, such as word vectors, based on included texts. The project has received attention, including open source contributions, and may become a successful digital humanities project in its own right.

Audiences

DH Box has been used in educational contexts, including workshops and classrooms, and has been featured at conferences via poster sessions and panel discussions.

The full reach of DH Box is difficult to quantify, as online usage data has not been collected due to privacy concerns. Nonetheless, we estimate that the number of DH Boxes created by users is in the thousands. To our knowledge, DH Box has been used in the US, Canada, and the UK.

Each of the workshops conducted by DH Box team members included groups ranging from 15 to 30 participants, representing a broad range of educational backgrounds (undergraduates and graduate students as well as faculty) and institutional backgrounds (Ivy Leagues, community colleges, urban and land grant universities). In particular, our institution is part of the CUNY system, one of the most ethnically and socioeconomically diverse universities in the country.

Workshops have been held at institutions including Columbia University, the CUNY Graduate Center, Hunter College, University of Victoria, the British Library, King's College London, and University of California Riverside. Conference participation with DH Box has occurred at the Canadian Society for Digital Humanities, University of Calgary; NYCDH Week, Fordham University; DH2017, McGill University; and MLA 2017 in Philadelphia. Currently, DH Box has 102 stars on GitHub, a metric that suggests engagement on that platform in the low thousands.

Evaluation

Considering the project in the context of the NEH Startup Grant, we feel DH Box has been a strong success. The team entered the grant period with a working prototype that has been dramatically improved and refined, most notably in the increased speed of environment creation, in the unified interface, in the addition of new tools, and in the deployment of a demonstration DH Box. The project has received widespread attention in the digital humanities community in addition to adoption and use at a number of institutions. It has also been generative as an example of infrastructure created by and for scholars in the humanities.

As a project, and especially as a humanities project, DH Box is technically sophisticated, with many moving parts including the web application, the interface to Docker, and the creation of "images" that are used as a template for environments. This is both an accomplishment and a potential drawback of the platform. DH Box performs a relatively difficult task—providing an

accessible interface to a discrete environment for digital work, an environment that might take hours to set up on a local machine. Creating this software with NEH funding has been a privilege and the culmination of a great deal of hard work and learning. However, given the relative complexity of the platform, the functions it performs, and the changing nature of the open source software that serves as its components, DH Box requires ongoing maintenance that might be a sustainability challenge in the future. Maintenance of the project in the medium- to long-term may be contingent on the team's ability to secure further funding for the project.

One potential avenue for DH Box was to run an instance that could be used as a service—that is, institutions could either set up their own DH Box using our one-line installer themselves, as they do now, or sign up for a DH Box service hosted centrally. The DH Box team has chosen not to pursue that route for a number of reasons. As a project largely driven by graduate students and faculty, the necessity of managing a full-time developer or team of developers would not be compatible with other duties, and DH Box was also never envisioned as a commercial or semi-commercial enterprise. In reality, however, most cloud-based projects, with a few exceptions such as OwnCloud, use the service model to some greater or lesser extent. While we stand by our decision not to offer DH Box as a service, this choice may constrain the adoption of DH Box as a platform, especially in institutions without resident technologists, and will affect the continuation and maintenance of DH Box in the medium- to long-term.

Continuation of the Project

DH Box has users, and in the short- to medium-term the project will be maintained by lead developer Stephen Zweibel. However, the environment in which DH Box exists in 2019 is different from the environment in 2013 when the project was conceived and 2014 when an initial version was created. In 2013 and 2014, cloud computing was in its nascency. The only way to teach tools such as Python or R was to install those tools locally—there were no commercial offerings allowing students to skip a lengthy process of downloading, installing, and configuring in a workshop setting. Since about 2016, however, offerings from startups and large corporations such as Google have entered this space. Services such as Glitch and REPL.io offer command line environments online, while platforms such as Code Ocean and Google's Colaboratory provide access to online notebooks similar to Jupyter Notebooks. These services are almost universally proprietary, and have bridged the infrastructural difficulties faced by DH Box (security, the necessity of monitoring for malicious activity, keeping component tools up-to-date) with the use of extensive financial, technical, and reputational resources. In 2015 and 2016, there was a moment when DH Box might have adapted an open source model supported by commercial activity, as do projects such as WordPress, Omeka, and Zotero. There may still be space for such a project, though the window may be closing. Though the DH Box team made the difficult choice not to adopt such a model, without it, it is unlikely for the project to carve out a niche in the medium- to long-term in an environment occupied by well-resourced corporations.

DH Box has created value for the DH community, for the Graduate Center, and for the DH Box team personally, and it is our hope that the project continues to be valuable in the coming years. In the medium- to long-term, however, the project may not prove to be actively maintainable without adoption of a model that provides greater resources over a longer period of time.

Long-Term Impact

DH Box has had a significant impact on the technical ecosystem at the Graduate Center, allowing Graduate Center Digital Initiatives to develop local technical expertise that has proven invaluable in securing and completing other major projects. Under the guidance of Matthew K. Gold, the Graduate Center maintains other pieces of scholarly infrastructure, most notably the Commons in a Box software used by institutions such as the MLA and various universities, but more recently the Mellon-funded Manifold Scholarship, a platform for publishing scholarly monographs on the web and in print. Jojo Karlin, outreach coordinator for DH Box, now serves as outreach coordinator for Manifold Scholarship. Stephen Zweibel, a graduate student at the Graduate Center in 2013, is now Digital Scholarship Librarian at the GC and oversees various DH-inflected projects. Patrick Smyth, who learned the Python programming language through working on DH Box, led technical curriculum development for a local intensive that served as the basis for a successful 2018-2019 Institutes for Advanced Topics in the Digital Humanities grant. Patrick was recently hired by Columbia University as Research Computing Program Coordinator. By funding DH Box, the NEH has developed local expertise that has facilitated the success of other projects, initiatives, and people at the Graduate Center even beyond the direct social and pedagogical value of DH Box for teaching courses and workshops.

Though more difficult to quantify, DH Box has strengthened relationships among the GC and other institutions and universities, most notably the University of Victoria and Columbia University but also including the New York Public Library, the British Library, King's College London, and Carleton University. From the beginning, the DH Box team was guided by faculty and staff at Columbia, including Dennis Tenen and Alexander Gil Fuentes. Jonathan Reeve was a DH Box team member from 2015, and Patrick Smyth now coordinates a research computing program based in Columbia University Libraries. Due to DH Box, representatives from the GC are now invited to lead a course at the Digital Humanities Summer Institute at the University of Victoria. Archeology Box, a fork of DH Box, is currently under development as a digital textbook at Carleton. The DH Box team has worked closely with these and other institutions to create instances of DH Box for use in technical workshops, or to answer questions about its use as the basis for future projects.

Though DH Box has built local infrastructure and relationships, and is a practical and usable piece of server-side software, the team also considers the project to be essential rhetorically, as an argument that humanities scholars can and should create their own infrastructure to address systemic problems in our own institutions. We understand that this approach is not a panacea or one that will apply in all (or even most) circumstances. However, as the academy becomes increasingly dominated by platforms created by well-financed private enterprises, creating our

own infrastructure as a means of carving out space for our own scholarly ethics becomes more important. For structural reasons, this avenue is not available to all. But those humanities scholars in the privileged position to create humanities infrastructure can provide others with tools that reflect their own, rather than technocratic, values.

The DH Box team would like to thank the National Endowment for the Humanities for supporting this project. Working on DH Box has allowed team members to grow as technologists, scholars, and advocates. We've had the opportunity to create a platform with social value while also contributing meaningfully to larger discussions about the digital humanities and infrastructure in the academy. The DH Box team looks forward to continuing to work on this and other DH-inflected projects after the close of the grant period.

Award Products

Conferences

"DH Box: A Push-Button Digital Humanities Laboratory"

METRO Conference

2015 — New York, NY

"Poster Session: DH Box"

Digital Humanities 2017

[Poster](#)

"DH Box"

[Open Source Lit, Open Source Crit panel](#)

MLA Convention

2017 - Philadelphia, PA

"DH Box: Building Cloud-Based Infrastructure for the Digital Humanities"

Canadian Society for Digital Humanities, Congress 2016

2016 - Calgary, Alberta

"DH Box Lunch Session"

Digital Humanities Summer Institute

2016 - Victoria, British Columbia

"DH Box: A Digital Humanities Laboratory in the Cloud"

CUNY DHI: Building a Digital Humanities Community

2015 — New York, New York

“DH Box: A Push-Button Digital Humanities Laboratory in the Cloud”

CUNY IT Conference

2014 — New York, New York

Workshops

Web APIs with Python

Digital Humanities Summer Institute, 2018 and 2019

Intro Python with DH Box

Interactive Technology and Pedagogy Program, Spring 2016

Getting Started with DH Box

CUNY Graduate Center Digital Fellows, Fall 2014

Publications and Press

[Tweet by Simon Tanner](#)

2017

[“The consequences of framing digital humanities tools as easy to use” \(Paige C Morgan\)](#)

College & Undergraduate Libraries, 2017

[“DH Box and Access in the Digital Humanities”](#)

dh+lib, 2017

[DH Box – Digital Humanities Virtual Workbench \(Tony Hirst\)](#)

2016

Software

[DH Box](#), a push-button digital humanities laboratory.

[Corpus](#), a textual corpus downloader for the digital humanities.

[DH-USB](#), a digital humanities operating system that runs on a USB drive.